

Influenza as a model system for studying the cross-species transfer and evolution of the SARS coronavirus

Robin M. Bush

Department of Ecology and Evolutionary Biology, 321 Steinhaus, University of California, Irvine, CA 92697, USA
(rmbush@uci.edu)

Severe acute respiratory syndrome coronavirus (SARS-CoV) moved into humans from a reservoir species and subsequently caused an epidemic in its new host. We know little about the processes that allowed the cross-species transfer of this previously unknown virus. I discuss what we have learned about the movement of viruses into humans from studies of influenza A, both how it crossed from birds to humans and how it subsequently evolved within the human population. Starting with a brief review of severe acute respiratory syndrome to highlight the kinds of problems we face in learning about this viral disease, I then turn to influenza A, focusing on three topics. First, I present a reanalysis of data used to test the hypothesis that swine served as a 'mixing vessel' or intermediate host in the transmission of avian influenza to humans during the 1918 'Spanish flu' pandemic. Second, I review studies of archived viruses from the three recent influenza pandemics. Third, I discuss current limitations in using molecular data to study the evolution of infectious disease. Although influenza A and SARS-CoV differ in many ways, our knowledge of influenza A may provide important clues about what limits or favours cross-species transfers and subsequent epidemics of newly emerging pathogens.

Keywords: SARS coronavirus; zoonosis; epidemic; pandemic; haemagglutinin

1. THE EMERGENCE OF SARS CORONAVIRUS

Modern molecular and analytical tools allowed the rapid identification of SARS-CoV, a new member of the coronavirus family (Drosten *et al.* 2003; Ksiazek *et al.* 2003; Peiris *et al.* 2003). Its genome was completely sequenced (Rota *et al.* 2003; Ruan *et al.* 2003) and the virus was confirmed as the cause of SARS (Fouchier *et al.* 2003) shortly after the start of the epidemic. Sequence data were subsequently used to track the spread and evolution of the virus (Zhong *et al.* 2003; He *et al.* 2004; Guan *et al.* 2004; Yeh *et al.* 2004). Although it is as yet unclear exactly when cross-species transfer occurred, the first human cases known are from late autumn of 2002, just months before the major outbreak in May 2003. Thus we may have isolated SARS-CoV very soon after its initial transmission to humans. This should surely lead to better information on host source than we have obtained for influenza A, where we are limited to the study of a few poorly preserved samples from decades past. Nonetheless, we as yet do not know where SARS-CoV came from.

Following up on reports that early cases of SARS occurred in animal handlers in the live markets of Guangdong Province, China, it was found that viruses very similar to human SARS-CoV could be isolated from masked palm civets (*Paguma larvata*) and raccoon dogs (*Nyctereutes procyonoides*), small rodent-eating mammals

native to Southeast Asia (Guan *et al.* 2003). This suggests that these animals served as intermediate hosts between a natural reservoir species and humans. However, efforts to isolate SARS-CoV from these species outside the live markets have failed.

SARS-CoV is relatively promiscuous: it has been shown to infect a wide range of mammals in the laboratory, including ferrets, domestic cats (Martina *et al.* 2003) and cynomolgus macaques (Rimmelzwaan *et al.* 2003). Thus the native host of SARS-CoV could be an unknown species that infected civets and other exotic food animals in their native habitat, on farms or *en route* to market. Rats have been suggested as agents of spread within the Amoy Hotel in Hong Kong, the primary epicentre of global spread (Ng 2003). Rodents might thus provide a common currency between the various types of small rodent-eating mammals found to harbour SARS-CoV in the markets. Unfortunately, as yet very little information is available on the occurrence of SARS-CoV in rodents in affected areas.

Despite the lack of definitive evidence that civets outside the market system pose a threat to human health, massive and controversial extermination campaigns against civets have subsequently been carried out. In part, this may have been inspired by the initially successful attempt to rid the Hong Kong markets of avian influenza in 1997. These avian influenza A subtype H5N1 viruses infected at least 18 humans, six of whom died (de Jong *et al.* 1997; Claas *et al.* 1998; Subbarao *et al.* 1998).

The H5N1 avian influenza A viruses responsible for the 1997 Hong Kong outbreak were unlike any known avian viruses. They appear, based on sequence data, to be

One contribution of 15 to a Discussion Meeting Issue 'Emerging infections: what have we learnt from SARS?'

reassortants between viruses from geese and viruses from either quail or teal (Guan *et al.* 1999; Hoffmann *et al.* 2000). These species are often caged in close proximity in live markets. Influenza viruses have a segmented genome and so are capable of forming reassortant progeny if two viruses infect a single host cell.

Unfortunately, repeated culling measures have failed to contain the problem permanently, and H5N1 viruses are currently causing a devastating pandemic in domestic fowl across Southeast Asia. Many humans coming in contact with these birds have contracted the virus; in some cases they have died. Fortunately, there is thus far no evidence that the H5N1 virus has become adapted for transmission between humans. Nonetheless, this outbreak is of great concern because the vast number of infections increases the probability of avian-human reassortment.

2. THE ORIGIN OF INFLUENZA A

One of the most interesting aspects of the 1997 H5N1 outbreak in Hong Kong is that prior to that time direct transmission of avian viruses to humans had been reported rarely and was believed to be highly restricted. Influenza viruses from humans and birds are known to bind preferentially to different forms of the sialic-acid receptor on host cells. This preferential binding was thought to be the primary barrier against human infection by avian strains and led to the idea that swine, whose cells possess both the receptors preferred by avian and human influenza viruses, serve as intermediate hosts or 'mixing vessels' for the transmission of avian viruses to humans (Scholtissek 1990).

This hypothesis is consistent with the observation that a massive outbreak of respiratory disease in swine occurred concurrently with the 1918 influenza pandemic in humans, and would explain why many epidemics and pandemics appear to originate in Southeast Asia, where agricultural practices put ducks, swine and humans in close contact, as reviewed by de Jong *et al.* (2000). Swine can clearly be infected by both human- and avian-adapted influenza viruses. However, the role of swine in the cross-species transfer of influenza A to humans is, despite much study, still unclear.

Here, I review two types of molecular analysis that have been used to try to determine the source of pandemic influenza viruses and the mechanisms by which they crossed species barriers. Both are, or probably will be, applied to the study of SARS; thus, I point out in some detail the limitations of these methods as well as what insight they can offer. At the end I review more general limitations in using molecular data to study the evolution of infectious disease.

3. RETROSPECTIVE ANALYSES BASED ON PHYLOGENETICS

One method for dating prior events is to use molecular data to estimate current rates of genetic change, and then extrapolate backwards in time to the period of interest. This method was used to test the hypothesis that swine served as a 'mixing vessel' for the reassortment of avian- and human-adapted influenza viruses in the origin of the 1918 'Spanish flu' pandemic (Scholtissek 1990).

Scholtissek *et al.* (1993b) constructed a phylogenetic tree using sequence data for the nucleoprotein genes of 23 human and 24 swine influenza viruses. They calculated the genetic distance from the root of the tree to each isolate, then regressed distance against isolation date to estimate an average rate of evolution in nucleotide substitutions per year. The resulting plot is redrawn in figure 1a. Assuming constant rates of evolution over time, they extrapolated backwards to the time (horizontal) axis to estimate when the original viruses were first transmitted to these new hosts. An estimate of the time of divergence from a common ancestor could have been obtained, had the lines not been parallel, from the point in time at which the lines crossed.

Figure 1a shows the swine lineage intercepting the time axis around 1912, slightly before the human lineage, which intercepts the line at around 1920. However, the authors noted that if they had displaced the root of the tree (which appears to have been rooted at the midpoint) 12 nucleotide substitutions nearer to the swine lineage, the human and swine influenza lineages would have both crossed the time axis at around 1918. Although the authors offer no definitive conclusions as to which new host was infected first, this analysis has often been used to suggest that an avian influenza virus was first transmitted to pigs and subsequently evolved the ability to infect humans around the time of the 1918 pandemic (Scholtissek *et al.* 1998; Webster 1998).

However, it is possible to move the root of a tree arbitrarily in any number of directions. In this example, moving the root across the possible rooting options (from the base of the swine clade to the base of the human clade) produces widely varying and contradictory conclusions. If the tree was rooted at the base of the human clade, it would appear that the virus first infected humans in 1899 and then swine 35 years later, in 1934. If the tree was rooted at the base of the swine clade, it would appear that the virus first infected swine in 1891 and then humans in 1942, 51 years later.

Obviously these rooting decisions should not be made arbitrarily. An outgroup should be used to root a tree if one is available. Adding A/Equine/Prague/56 (Reid *et al.* 2003) to the analysis shown in figure 1a suggests that the root should be moved four nucleotide substitutions closer to the swine lineage. Doing so suggests transmission to humans in 1900 and to swine in 1922, dates that are inconsistent with observed disease incidence. The use of a different outgroup sequence could well give different results.

Unfortunately, for many groups of organisms the outgroup is unknown or may be only distantly related to the lineages of interest. This point is especially germane to the study of SARS-CoV because determining its nearest relative has proved problematic (Drosten *et al.* 2003; Eickmann *et al.* 2003; Marra *et al.* 2003; Rota *et al.* 2003) and may never be resolved.

Another major limitation to these types of regression analyses is that they are very sensitive to the particular dataset used, especially when sample sizes are small. In this example, Scholtissek *et al.* (1993b) employed only 77% of the data points used to construct the phylogeny when estimating the regression lines in figure 1a. The stated exclusion criterion was that the excluded points lay

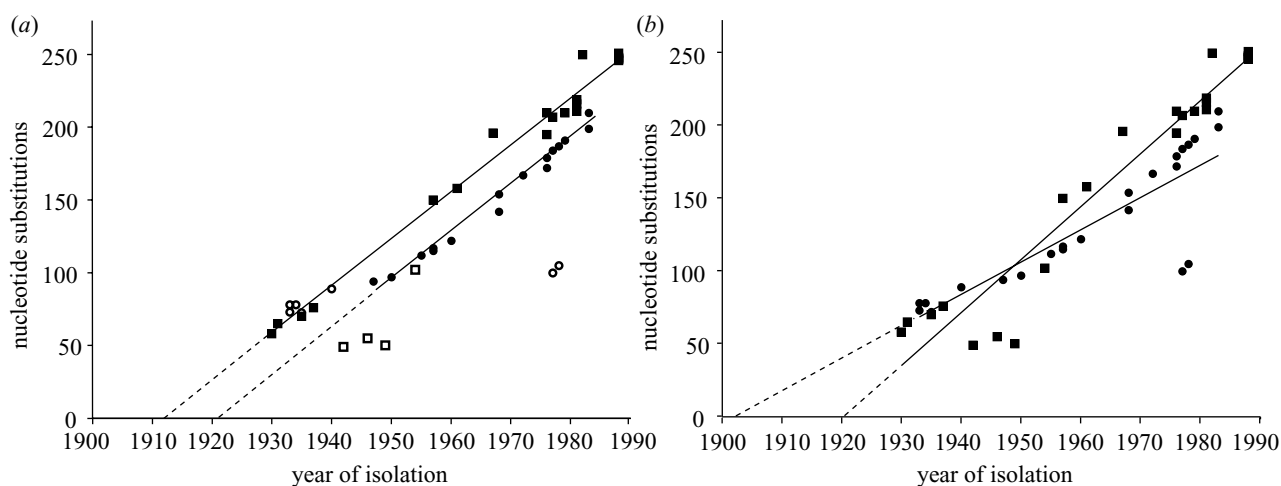


Figure 1. Cross-species transmission estimates for human and swine influenza A subtype H1N1. Genetic distances are measured from the root of a phylogenetic tree (not shown). Data from Scholtissek *et al.* (1993*a,b*). Rates of evolution for swine (squares) and human (circles) lineages are calculated as the slopes of the respective regression lines. Units are nucleotide substitutions per year. Extrapolation backwards in time (dashed lines) was used to determine the date of the initial transmission of the virus into these hosts. (a) Rates estimated using only the closed symbols, as in Scholtissek *et al.* (1993*b*). (b) Rates estimated using the complete dataset.

too far from the regression lines (Scholtissek *et al.* 1993*a*); the criteria for establishing those lines in the first place were not provided. Figure 1*b* shows the resulting plot had all data points been included. These results suggest that transmission to humans occurred in 1904 and to swine in 1921 (figure 1*b*). Neither of these dates are consistent with historical observations of disease incidence. In addition, the two regression lines diverge rather than converge as they approach the time axis. This gives the impression that these lineages did not diverge from a common ancestor; however, both are believed to have originated from avian strains (Reid *et al.* 2003). Clearly, this dataset would have provided no support for the 'mixing vessel' hypothesis if all the data used to construct the phylogeny had also been used in the regression analysis.

There is always a risk in drawing conclusions from extrapolation of a regression analysis (Kuo 2002). In the case of emerging infectious disease, this technique is especially suspect because extrapolation relies on the assumption of a constant rate of evolution over time. The 1918 pandemic infected humans in waves of increasing severity in 1918 and 1919 before evolving into the (relatively) benign form we experience today. To assume a constant rate of evolution over this entire period is questionable.

As noted by Cox *et al.* (1993) the influenza literature reports substantial variation in the rates of evolution for the different strains, even during very recent periods of time when the initial adaptation to humans is presumably over. One major cause of this variation is lack of data, another is drawing conclusions using data that cover only short periods of time. An illustration of rate variation for influenza A subtype H3N2 is shown in figure 2*a*.

Varying estimates of evolutionary rates have already been reported for SARS-CoV (He *et al.* 2004; Yeh *et al.* 2004) despite the very short period of time it has been under study. Based on our experience with influenza, these estimates will change not only over time, if the virus continues to circulate, but also with the addition of more data for the time periods already studied.

4. ANALYSIS OF ARCHIVED INFLUENZA VIRUSES

The origins of pandemic influenza have also been examined through the study of archived viruses. The pandemics of 1957 and 1968 were clearly caused by reassortant viruses containing human and avian influenza genes. The influenza A genome is composed of eight segments, each containing one or two of its 10 genes. Influenza strains are typically referred to by the genetic variants of their surface proteins, haemagglutinin and neuraminidase. At present 15 haemagglutinin alleles (numbered H1–H15) and nine neuraminidase alleles (N1–N9) are known from waterfowl. These avian viruses are thought to be the ancestors of strains currently circulating in swine, horses and humans (Webster *et al.* 1992).

The 1918 pandemic strain carried H1 and N1 alleles. A descendant of this strain appears to have gained avian-derived genes for surface proteins H2 and N2, and for PB1, one of the influenza polymerase genes, through reassortment in 1957. The resulting H2N2 strain circulated in humans until 1968 when it was replaced by a reassortant containing new avian H3 and PB1 genes (Scholtissek *et al.* 1978; Kawaoka *et al.* 1989). The resulting H3N2 virus continues to circulate in humans today. Although these reassortment events may have taken place within swine, there is no evidence to support this thesis from the sequence data, which implicate only avian and human sources.

The origin of the deadly 1918 pandemic is less clear than those of the 1957 and 1968 pandemics. Ongoing studies of H1N1 influenza A viruses preserved in the archived lung tissue of two army soldiers and from an Alaskan Inuit woman frozen in permafrost, all victims of the 1918 pandemic, have yet to reveal why this strain was so deadly or exactly where it came from (Taubenberger *et al.* 1997; Reid *et al.* 1999). The haemagglutinin and neuraminidase alleles resemble the oldest available classical H1N1 swine influenza strains (from 1930), but share characteristics with modern avian H1N1 strains as well. Sequencing viruses isolated from waterfowl collected in

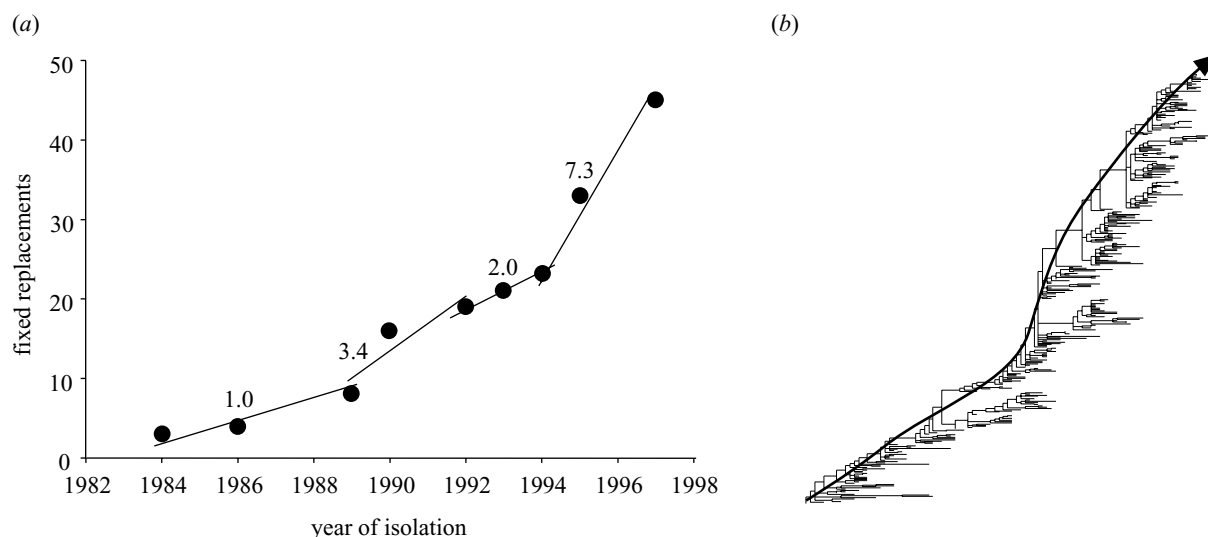


Figure 2. Variation over time in the rate of influenza A subtype H3N2 evolution. (a) Circles show the cumulative number of amino acid replacements fixed along the trunk of the tree (see arrow in (b); data from Bush *et al.* (1999)). The numbers indicate rates per year calculated over arbitrarily chosen short intervals of time as indicated by regression lines. Choosing different intervals would result in vastly different rate estimates.

1917 and preserved in alcohol in the American Museum of Natural History has done little to resolve this mystery (Fanning *et al.* 2002; Reid *et al.* 2003). An additional line of inquiry stems from X-ray crystallographic studies of haemagglutinin proteins reconstructed from 1918 sequence data. These data suggest that the binding site of the 1918 human virus is more avian-like than that of later H1N1 viruses (Gamblin *et al.* 2004; Stevens *et al.* 2004). But, as Reid *et al.* (2003) concluded, it appears, based on current material, that, if the 1918 human pandemic strain was avian-derived, it must have evolved undetected in a non-avian host for some time prior to the 1918 human pandemic.

We have some knowledge of the molecular basis of host specificity for influenza viruses, such as the presence or absence of a sequence of basic amino acids at the haemagglutinin cleavage site, and a preferential binding to the α 2,3-linked galactosidase found in birds rather than the α 2,6-linked galactosidase found on human lung cells (reviewed by Zambon 2004). However, binding studies clearly showed these differences to be preferences rather than absolute barriers to infection (Matrosovich *et al.* 1993). This result is sadly supported by the many recent infections of humans with entirely avian viruses.

Although it has long been known that sporadic infections of humans by avian viruses can occur (Shortridge 1992), transmission within the new host population is rare. Efficient transmission seemingly depends on a number of variables, and may well require that interacting coadapted sets of genes remain together through reassortment events (Rott 1992). New experiments using reverse genetics to construct influenza viruses with various combinations of human and avian genes will hopefully provide greater insight into the genetics of host specificity and modes of transmission. (Neumann *et al.* 2003).

Evidence for direct infection of humans by avian viruses does not prove that swine have never been involved in the transmission of avian influenza to humans. It suggests, however, the existence of additional barriers to establishment in mammals. One barrier may be the lack of efficient

transmission between individuals in the new host species. Birds generally harbour influenza viruses in their intestinal tract, not in their lungs. Thus avian viruses must adapt both to conditions in the mammalian respiratory tract and to airborne transmission. Dehydration during aerosol transmission among humans, for example, is a challenge not experienced during spread in faeces or in the aquatic environments of waterfowl. Differences in temperature and pH may also play a role.

The genetics of transmission is clearly an area in need of study, but by its nature it is an impossible problem to address using humans. Although cynomolgus macaques infected with avian H5N1 influenza A produced a necrotizing pneumonia similar to that seen in the human fatalities of H5N1 infection (Rimmelzwaan *et al.* 2003), studying transmission using these animals is formidably expensive and in some eyes unethical, and, in addition, there is no guarantee that the results would be applicable to humans. Transmission studies of SARS-CoV in animal models might be similarly expensive and difficult to interpret.

5. LIMITATIONS OF MOLECULAR DATA

The existing influenza sequence data are among the best available for studying the evolution of infectious disease. However, there are problems with using these data to study influenza evolution and population biology, and these limitations may hold true for SARS-CoV as well. One problem is the presence of laboratory artefacts in the sequence data. Although cell culture is increasingly used, amplification of the influenza virus by passage in embryonated hens' eggs has been standard laboratory practice for the culture of influenza viruses for many years. Egg passage is still required for strains that will be used in the influenza vaccine in the USA. Unfortunately, the haemagglutinin of human influenza viruses evolves rapidly to adapt to replication in eggs (Robertson 1993). The resulting sequences may thus contain replacements that either were not present or were at low frequency in the

original viral sample. These laboratory artefacts often occur at sites involved in adaptation to humans as well as to eggs (Cox & Bender 1995).

It is possible to estimate the proportion of amino acid replacements resulting from egg passage by comparing the numbers of replacements found in sequences in cell-passaged and egg-passaged isolates (Bush *et al.* 2000, 2001). In the HA1 domain of influenza A subtype H3N2 haemagglutinin, egg passage was associated with *ca.* 8% of amino acid replacements (Bush *et al.* 1999). Unfortunately, in the absence of controls—viruses that have never been passaged—it is impossible to determine which replacements in a dataset are artefacts.

These artefacts inflate the amount of evolutionary change that one infers from sequence data. Because these artefacts are non-synonymous as opposed to synonymous substitutions, care must be taken to eliminate them from analyses seeking evidence of positive selection by the human immune system. One way to minimize such error is to discard changes assigned to the terminal branches of the trees when estimating substitution rates (Bush *et al.* 1999). Studies of positive selection in influenza that fail to exclude replacements selected for during egg passage routinely find evidence for selection on codons for which there is no evidence of a selective advantage in humans (Yang 2000; Yang *et al.* 2000; Huelsenbeck *et al.* 2001; Nielsen & Huelsenbeck 2002). In studies of positive selection thus far in SARS-CoV, some groups deleted possible artefacts (Ruan *et al.* 2003; He *et al.* 2004), while Yeh *et al.* (2004), after contrasting a direct PCR product with sequences from isolates cultured in monkey kidney cells, did not find culture-induced artefacts to be a problem. These studies have so far found variable evidence for positive selection in SARS-CoV, which is not surprising given how few data are as yet available.

Another difficulty in the molecular analysis of sequence data collected during disease surveillance is sampling bias. The WHO influenza surveillance system is purposefully biased towards sequencing viruses that differ antigenically from commonly sampled strains on the basis of the haemagglutination inhibition test. This bias causes an overestimation of positive selection on the haemagglutinin gene because only non-synonymous substitutions produce antigenic change. The WHO is the main source of influenza sequence data; thus this sampling bias is reflected in the composition of sequences present in GenBank. Assuming that the frequencies of various genetic groups in GenBank reflect their frequencies in nature (Plotkin *et al.* 2002) will invariably lead to erroneous results under current WHO sampling protocols.

Last, it can be very difficult to make accurate inferences about evolutionary relationships between distantly related organisms because of the resulting sequence dissimilarity. Conclusions may vary dramatically depending on how these sequences are aligned. Early reports that some genes in the SARS-CoV genome are the result of recombination (Rest & Mindell 2003; Stavriniades & Guttman 2004) may be alignment dependent. They may also share characteristics with a study claiming that the 1918 influenza haemagglutinin gene was a recombinant (Gibbs *et al.* 2001). This study has been criticized for not being robust with respect to the method of phylogenetic reconstruction (Worobey *et al.* 2002). Ideas about the recombinant origin

of SARS-CoV may well change as more data become available.

6. SUGGESTIONS FOR FUTURE RESEARCH

The extents of the spread of most infectious diseases are vastly understudied in part because there is almost no emphasis on determining the occurrence of subclinical disease. Farmers in Southeast Asia have long been reported to carry antibodies to a number of avian influenza subtypes not known to circulate in humans, including the H5 allele, which was recently involved in outbreaks of human illness in Hong Kong (Shortridge 1992). Sera from healthy blood donors in Hong Kong contained antibodies to the H9N2 virus, suggesting prior infection by this strain (Peiris *et al.* 1999). Early serological reports suggested a subclinical infection rate of 13% in animal traders (CDC 2003); however, as we learn more about the serological cross-reactivity of SARS-CoV with common coronaviruses such values may change. Surveillance rarely targets healthy people or geographical locations not experiencing a high incidence of disease. This may be why we are so often surprised by new outbreaks of infectious disease.

We may also continue to be surprised if we expect new epidemics to arise from viruses that evolve from the most recently circulating strains. This is not always the case: in many instances new influenza-epidemic strains are descendants of viruses from years past, viruses that had persisted at low frequency while other strains caused our yearly epidemics (Cox *et al.* 1993). Because extensive surveillance for influenza has been in place for over 50 years, the influenza surveillance community is often aware of these lurking threats. Unfortunately, global surveillance does not exist for most known pathogens and is certainly lacking for those viruses, like SARS-CoV, that have yet to emerge from their even more poorly known animal hosts. Funding for such efforts is discussed in the heat of an outbreak; however, effective surveillance, even of human infectious diseases, is a long way from becoming a reality. Even less interest and money is being directed towards conservation of museum and medical archives, which as discussed in § 4 have contributed much of what we know about the origin of pandemic influenza. One wonders whether tissue samples are being saved from the masked palm civets currently being destroyed in China: we may be in the process of burning the evidence.

REFERENCES

- Bush, R. M., Fitch, W. M., Bender, C. A. & Cox, N. J. 1999 Positive selection on the H3 hemagglutinin gene of human influenza virus A. *Mol. Biol. Evol.* **16**, 1457–1465.
- Bush, R. M., Smith, C. B., Cox, N. J. & Fitch, W. M. 2000 Effects of passage history and sampling bias on phylogenetic reconstruction of human influenza A evolution. *Proc. Natl Acad. Sci. USA* **97**, 6974–6980.
- Bush, R. M., Fitch, W. M., Smith, C. B. & Cox, N. J. 2001 Predicting influenza evolution: the impact of terminal and egg-adapted mutations. In *Options for the control of influenza*, vol. 4 (ed. A. D. M. E. Osterhaus), pp. 147–153. Amsterdam: Elsevier.
- CDC 2003 Prevalence of IgG antibody to SARS-associated coronavirus in animal traders: Guangdong Province, China, 2003. *Morbidity Mortality Wkly Rep.* **17**, 986–987.

- Claas, E. C., Osterhaus, A. D., van Beek, R., De Jong, J. C., Rimmelzwaan, G. F., Senne, D. A., Krauss, S., Shortridge, K. F. & Webster, R. G. 1998 Human influenza A H5N1 virus related to a highly pathogenic avian influenza virus. *Lancet* **351**, 472–477.
- Cox, N. J. & Bender, C. A. 1995 The molecular epidemiology of influenza viruses. *Seminars Virol.* **6**, 359–370.
- Cox, N., Xu, X., Bender, C., Kendal, A., Regnery, H., Hemphill, M. & Rota, P. 1993 Evolution of hemagglutinin in epidemic variants and selection of vaccine viruses. In *Options for the control of influenza*, vol. 2 (ed. C. Hannoun, A. P. Kendal, H. D. Klenk & F. L. Ruben), pp. 223–230. Amsterdam: Elsevier.
- de Jong, J. C., Claas, E. C., Osterhaus, A. D., Webster, R. G. & Lim, W. L. 1997 A pandemic warning? *Nature* **389**, 554.
- de Jong, J. C., Rimmelzwaan, G. F., Fouchier, R. A. M. & Osterhaus, A. 2000 Influenza virus: a master of metamorphosis. *J. Infect.* **40**, 218–228.
- Drosten, C. (and 25 others) 2003 Identification of a novel coronavirus in patients with severe acute respiratory syndrome. *New Engl. J. Med.* **348**, 1967–1976.
- Eickmann, M. (and 15 others) 2003 Phylogeny of the SARS coronavirus. *Science* **302**, 1504–1505.
- Fanning, T. G., Slemons, R. D., Reid, A. H., Janczewski, T. A., Dean, J. & Taubenberger, J. K. 2002 1917 avian influenza virus sequences suggest that the 1918 pandemic virus did not acquire its hemagglutinin directly from birds. *J. Virol.* **76**, 7860–7862.
- Fouchier, R. A., Kuiken, T., Schutten, M., Van Amerongen, G., Van Doornum, G. J., Van den Hoogen, B. G., Peiris, M., Lim, W., Stohr, K. & Osterhaus, A. D. 2003 Aetiology: Koch's postulates fulfilled for SARS virus. *Nature* **423**, 240.
- Gamblin, S. J. (and 11 others) 2004 The structure and receptor-binding properties of the 1918 influenza hemagglutinin. *Science* **303**, 1838–1842.
- Gibbs, M. J., Armstrong, J. S. & Gibbs, A. J. 2001 Recombination in the hemagglutinin gene of the 1918 'Spanish flu'. *Science* **293**, 1842–1845.
- Guan, Y., Shortridge, K. F., Krauss, S. & Webster, R. G. 1999 Molecular characterization of H9N2 influenza viruses: were they the donors of the 'internal' genes of H5N1 viruses in Hong Kong? *Proc. Natl Acad. Sci. USA* **96**, 9363–9367.
- Guan, Y. (and 17 others) 2003 Isolation and characterization of viruses related to the SARS coronavirus from animals in southern China. *Science* **302**, 276–278.
- Guan, Y. (and 17 others) 2004 Molecular epidemiology of the novel coronavirus that causes severe acute respiratory syndrome. *Lancet* **363**, 99–104.
- He, J. F. (and 52 others) 2004 Molecular evolution of the SARS coronavirus during the course of the SARS epidemic in China. *Science* **303**, 1666–1669.
- Hoffmann, E., Stech, J., Leneva, I., Krauss, S., Scholtissek, C., Chin, P. S., Peiris, M., Shortridge, K. F. & Webster, R. G. 2000 Characterization of the influenza A virus gene pool in avian species in southern China: was H6N1 a derivative or a precursor of H5N1? *J. Virol.* **74**, 6309–6315.
- Huelsenbeck, J. P., Ronquist, F., Nielsen, R. & Bollback, J. P. 2001 Bayesian inference of phylogeny and its impact on evolutionary biology. *Science* **294**, 2310–2314.
- Kawaoka, Y., Krauss, S. & Webster, R. G. 1989 Avian-to-human transmission of the PB1 gene of influenza A viruses in the 1957 and 1968 pandemics. *J. Virol.* **63**, 4603–4608.
- Ksiazek, T. G. (and 25 others) 2003 A novel coronavirus associated with severe acute respiratory syndrome. *New Engl. J. Med.* **348**, 1953–1966.
- Kuo, Y.-H. 2002 Extrapolation of correlation between 2 variables in 4 general medical journals. *JAMA* **287**, 2815–2817.
- Marra, M. A. (and 58 others) 2003 The genome sequence of the SARS-associated coronavirus. *Science* **300**, 1399–1404.
- Martina, B. E., Haagmans, B. L., Kuiken, T., Fouchier, R. A., Rimmelzwaan, G. F., Van Amerongen, G., Peiris, J. S., Lim, W. & Osterhaus, A. D. 2003 Virology: SARS virus infection of cats and ferrets. *Nature* **425**, 915.
- Matrosovich, M. N., Gambaryan, A. S., Tuzikov, A. B., Byaromova, N. E., Mochalova, L. V., Golbraikh, A. A., Shenderovich, M. D., Finne, J. & Bovin, N. V. 1993 Probing of the receptor-binding sites of the H1 and H3 influenza A and influenza B virus hemagglutinins by synthetic and natural sialosides. *Virology* **196**, 111–121.
- Neumann, G., Hatta, M. & Kawaoka, Y. 2003 Reverse genetics for the control of avian influenza. *Avian Dis.* **47**, 882–887.
- Ng, S. K. 2003 Possible role of an animal vector in the SARS outbreak at Amoy Gardens. *Lancet* **362**, 570–572.
- Nielsen, R. & Huelsenbeck, J. P. 2002 Detecting positively selected amino acid sites using posterior predictive P-values. In *Pacific Symposium on Biocomputing* (ed. R. B. Altman, A. K. Dunker, L. Hunter, K. Lauderdale & T. E. Klein), pp. 576–588. River Edge, NJ: World Scientific.
- Peiris, J. S. (and 15 others) 2003 Coronavirus as a possible cause of severe acute respiratory syndrome. *Lancet* **361**, 1319–1325.
- Peiris, M., Yuen, K. Y., Leung, C. W., Chan, K. H., Ip, P. L., Lai, R. W., Orr, W. K. & Shortridge, K. F. 1999 Human infection with influenza H9N2. *Lancet* **354**, 916–917.
- Plotkin, J. B., Dushoff, J. & Levin, S. A. 2002 Hemagglutinin sequence clusters and the antigenic evolution of influenza A virus. *Proc. Natl Acad. Sci. USA* **99**, 6263–6268.
- Reid, A. H., Fanning, T. G., Hultin, J. V. & Taubenberger, J. K. 1999 Origin and evolution of the 1918 'Spanish' influenza virus hemagglutinin gene. *Proc. Natl Acad. Sci. USA* **96**, 1651–1656.
- Reid, A. H., Fanning, T. G., Slemons, R. D., Janczewski, T. A., Dean, J. & Taubenberger, J. K. 2003 Relationship of pre-1918 avian influenza HA and NP sequences to subsequent avian influenza strains. *Avian Dis.* **47**, 921–925.
- Rest, J. S. & Mindell, D. P. 2003 SARS associated coronavirus has a recombinant polymerase and coronaviruses have a history of host-shifting. *Infect. Genet. Evol.* **3**, 219–225.
- Rimmelzwaan, G. F., Kuiken, T., Van Amerongen, G., Bestebroer, T. M., Fouchier, R. A. & Osterhaus, A. D. 2003 A primate model to study the pathogenesis of influenza A (H5N1) virus infection. *Avian Dis.* **47**, 931–933.
- Robertson, J. S. 1993 Clinical influenza virus and the embryonated hen's egg. *Rev. Med. Virol.* **3**, 97–106.
- Rota, P. A. (and 34 others) 2003 Characterization of a novel coronavirus associated with severe acute respiratory syndrome. *Science* **300**, 1394–1399.
- Rott, R. 1992 The pathogenic determinant of influenza virus. *Vet. Microbiol.* **33**, 303–310.
- Ruan, Y. J. (and 19 others) 2003 Comparative full-length genome sequence analysis of 14 SARS coronavirus isolates and common mutations associated with putative origins of infection. *Lancet* **361**, 1779–1785.
- Scholtissek, C. 1990 Pigs as 'mixing vessels' for the creation of new pandemic influenza A viruses. *Med. Principles Practice* **2**, 65–71.
- Scholtissek, C., Rohde, W., Von Hoyningen, V. & Rott, R. 1978 On the origin of the human influenza virus subtypes H2N2 and H3N2. *Virology* **87**, 13–20.
- Scholtissek, C., Ludwig, S. & Fitch, W. M. 1993a Analysis of influenza-A virus nucleoproteins for the assessment of molecular genetic mechanisms leading to new phylogenetic virus lineages. *Arch. Virol.* **131**, 237–250.

- Scholtissek, C., Schultz, U., Ludwig, S. & Fitch, W. M. 1993b The role of swine in the origin of pandemic influenza. In *Options for the control of influenza*, II (ed. C. Hannoun, A. P. Kendal, H. D. Klenk & F. L. Ruben), pp. 193–201. Amsterdam: Elsevier.
- Scholtissek, C., Hinshaw, V. S. & Olsen, C. W. 1998 Influenza in pigs and their role as the intermediate host. In *Textbook of influenza* (ed. K. G. Nicholson, R. G. Webster & A. J. Hay), pp. 137–145. Oxford: Blackwell Science.
- Shortridge, K. F. 1992 Pandemic influenza: a zoonosis? *Seminars Respiratory Infect.* **7**, 11–25.
- Stavrinos, J. & Guttman, D. S. 2004 Mosaic evolution of the severe acute respiratory syndrome coronavirus. *J. Virol.* **78**, 76–82.
- Stevens, J., Corper, A. L., Basler, C. F., Taubenberger, J. K., Palese, P. & Wilson, I. A. 2004 Structure of the uncleaved human H1 hemagglutinin from the extinct 1918 influenza virus. *Science* **303**, 1866–1870.
- Subbarao, K. (and 15 others) 1998 Characterization of an avian influenza A (H5N1) virus isolated from a child with a fatal respiratory illness. *Science* **279**, 393–396.
- Taubenberger, J. K., Reid, A. H., Krafft, A. E., Bijwaard, K. E. & Fanning, T. G. 1997 Initial genetic characterization of the 1918 ‘Spanish’ influenza virus. *Science* **275**, 1793–1796.
- Webster, R. G. 1998 Evolution and ecology of influenza viruses: interspecies transmission. In *Textbook of influenza* (ed. K. G. Nicholson, R. G. Webster & A. J. Hay), pp. 109–119. Oxford: Blackwell Science.
- Webster, R. G., Bean, W. J., Gorman, O. T., Chambers, T. M. & Kawaoka, Y. 1992 Evolution and ecology of influenza A viruses. *Microbiol. Rev.* **56**, 152–179.
- Worobey, M., Rambaut, A., Pybus, O. G. & Robertson, D. L. 2002 Questioning the evidence for genetic recombination in the 1918 ‘Spanish flu’ virus. *Science* **296**, 211.
- Yang, Z. 2000 Maximum likelihood estimation on large phylogenies and analysis of adaptive evolution in human influenza virus A. *J. Mol. Evol.* **51**, 423–432.
- Yang, Z., Nielsen, R., Goldman, N. & Pedersen, A. M. 2000 Codon-substitution models for heterogeneous selection pressure at amino acid sites. *Genetics* **155**, 431–449.
- Yeh, S. H. (and 19 others) 2004 Characterization of severe acute respiratory syndrome coronavirus genomes in Taiwan: molecular epidemiology and genome evolution. *Proc. Natl Acad. Sci. USA* **101**, 2542–2547.
- Zambon, M. 2004 The inexact science of influenza prediction. *Lancet* **363**, 582–583.
- Zhong, N. S. (and 15 others) 2003 Epidemiology and cause of severe acute respiratory syndrome (SARS) in Guangdong, People’s Republic of China, in February 2003. *Lancet* **362**, 1353–1358.

GLOSSARY

- SARS: severe acute respiratory syndrome
 SARS-CoV: severe acute respiratory syndrome coronavirus
 WHO: World Health Organization